

Institute of Geography Online Paper Series: GEO-015

Rural and Urban Road Network Generalisation

Deriving 1:250,000 from OS MasterMap

Omar Chaudhry & William Mackaness

Institute of Geography, School of GeoSciences,
University of Edinburgh,
Drummond St,
Edinburgh EH8 9XP
O.Chaudhry@sms.ed.ac.uk
William.mackaness@ed.ac.uk

XXII International Cartographic Conference (ICC2005)



Copyright

This online paper may be cited in line with the usual academic conventions. You may also download it for your own personal use. This paper must not be published elsewhere (e.g. mailing lists, bulletin boards etc.) without the author's explicit permission

Please note that :

- it is a draft;
- this paper should not be used for commercial purposes or gain;
- you should observe the conventions of academic citation in a version of the following or similar form:

Omair Chaudhry & William Mackaness (2005 *Rural and Urban Road Network Generalisation Deriving 1:250,000 from OS MasterMap*, online papers archived by the Institute of Geography, School of Geosciences, University of Edinburgh.

Rural and Urban Road Network Generalisation Deriving 1:250,000 from OS MasterMap

Omaisr Chaudhry & William Mackaness

Institute of Geography, School of GeoSciences, University of Edinburgh, Drummond St, Edinburgh EH8 9XP
O.Chaudhry@sms.ed.ac.uk, William.mackaness@ed.ac.uk

ABSTRACT:

Roads are essential component of topographic maps and spatial databases. The challenge in automated generalisation of road networks is to derive a connected network while maintaining the structure for the intended target scale and to achieve this with minimum user intervention. A lot of methods to select, displace and simplify roads have been presented; the focus here is on the generalisation of networks using visual perception techniques. This paper presents a framework based on visual perception that uses minimum attributes for generalisation of both 'rural' and 'urban' roads over large scale change. The system incorporated graph theoretic techniques to explicitly model the topology of the network as it was generalized. The model uses a fine scale map (1:1250 or 1:2500) as input and generates small scale (1:250,000) maps directly from it without creating intermediate small scale maps. The results compared favorably with paper maps (Ordnance Surveys Stratgie dataset (1:250,000)).

Keywords: Roads, Graph, Visual Perception, Strokes, Generalisation

1. INTRODUCTION:

Road networks are an essential component of maps – providing important contextual information, critical to the interpretation of human presence and activities. A lot of research has been devoted to road network generalisation in past decades. This is due to the fact that generalisation of linear/network map objects (such as roads, rivers, and railroads) is quite complex for a number of reasons. Firstly such objects are always connected to each other to form a coherent network. Deletion of roads without consideration of the entire network will result in disconnections. Secondly these objects have a certain shape, angle, orientation and size (length) instead of just a location and a value of importance. All these factors make the generalisation of roads a non trivial process.

This paper presents a framework based on visual perception that uses minimum attributes for generalisation of both rural and urban roads over large scale change. Instead of the traditional incremental approach, the focus is on generation of very small scale mapping directly from fine scale without creating intermediate results. This requires a design capable of handling large volumes of data, one that is also able to deal with the sparse network of rural areas, and the dense networks associated with built up areas. The system was based on the idea proposed by (Thomson and Richardson, 1999; Zhang, 2004) according to which road segments that are in continuation or seem to be in continuation visually are aggregated/merged together resulting in 'strokes' (a chain of road segments). The system incorporates graph theoretic techniques to explicitly model the topology of the network as it is generalized.

The paper is organized as follows: Section 2 gives a brief introduction to perceptual visualization principles; Section 3 details the proposed methodology and explains the implementation using a small test data. Section 4 presents a case study, Section 5 evaluates the results followed by areas of further improvement. We conclude with a summary.

2.0 BACKGROUND:

Before going into the actual design and implementation it is important to provide a brief introduction to perceptual organization and especially the 'Principle of Good Continuation'. These concepts are vital in understanding of the proposed methodology and implementation.

2.1 Perceptual Organization:

The importance of perceptual organization in map generalisation has long been recognized. Notably DeLucia and Black (1987 p.175) stated "... there are direct analogies between what is required for successful map generalization and ...

the principles of perceptual organization enunciated long ago by the Gestalt psychologists”. Perceptual Organization is part of visual perception which in simple terms is the identification, organization, and interpretation of sensory data received by the individual through the eye. The Gestalt psychologists formulated a number of principles of perceptual organization to describe how certain perceptions are more likely to occur than others (Bruce et al., 2003). They argued that when we view the world we do not see a collection of edges and blobs but instead we see an organized world of surfaces and objects. Many perceptual principles have been identified: the main ones being proximity, similarity, good continuation, closure, smallness, surrounded ness, symmetry.

In making a visual comparison between source data (OS MasterMap) with the target scale map (Strategi 1:250,000). It was noticed that one tends to combine different roads in order of increasing salience. This principle is termed ‘the law of good continuation’. The principle of good continuation states that objects arranged in either a straight line or a smooth curve tends to be seen as a unit. Thus when two long straight roads cross each other, we see them as two intersecting roads not as four separate lines meeting at a common intersection.

The experiment conducted by Field et al., (1993) provides a better understanding of the principle of good continuation as illustrated in Figure 1. In Figure 1A the observer’s task was to determine which of two such arrays contained ‘snake’. The experiments varied the ‘wiggleness’ of the snake (Figure 1B), and the degree to which the path elements were aligned or misaligned along the direction of the path (Figure 1C). Straight path, with elements aligned along the path were most easily detected. From the experiments was found out that observers combined the objects together with angle change from 40° to 60°. Field et al., (1993) interpret these findings in terms of an “association field”, suggesting “a localized linking process or association between the responses to the elements in the path according to a specific set of rules” (p.185). Their scheme of linking the process is shown in Figure 1D, indicating that links will be made between two adjacent stimulus elements if their orientation and position are such that they would lie on simple smooth curve passing through both elements. A more detail on their experiment can be found in ‘Visual Perception’ (Bruce et al., 2003). This idea was used in generalisation of roads in this paper.

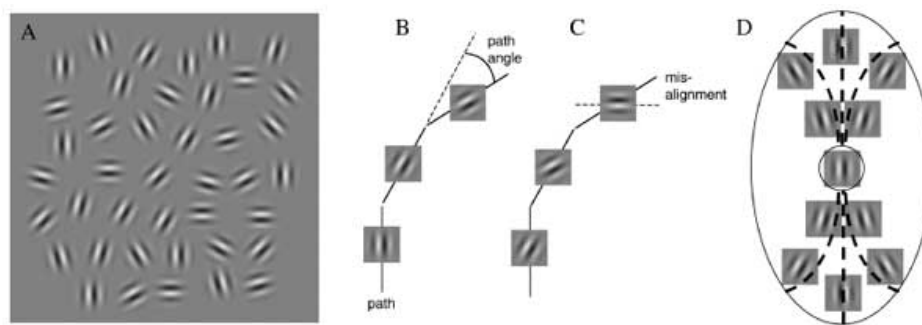


Figure 1: An example of the experiment conducted by Field, Hayes and Hess (1993) for understanding the principle of good continuation (Bruce et al., 2003)

3.0 METHODOLOGY:

The entire generalisation process was comprised of five stages (Figure 2). Stage I and II enriched the database and formed the basis for road selection (III). Stage IV checked connectivity among the selected roads. In the following sections each stage is discussed along with the test data to show the output at each stage: For complete discussion on the implementation see Chaudhry, 2004).

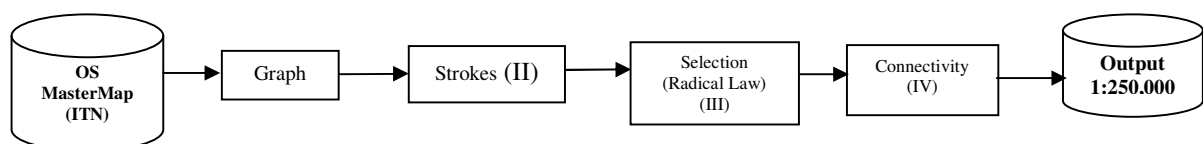


Figure 2: Summary of key stages in the generalization of road networks.

3.1 Data Analysis:

The first step was to analyze the input data set. The input data set used was Ordnance Survey (OS) Integrated Transport Network (ITN) which is comprehensive and is a detailed representation of the public and private road network of UK (Ordnance Survey, 2004). The data was converted into tabular format from shape file and loaded into Oracle. Java Data Base Connectivity (JDBC) was used to establish communication between Oracle and Java (Sun, 2004). The SQL package in Java was used to manipulate with the objects between the database and program. Figure 3 shows the test data after being loaded into Oracle.



Figure 3: A part of actual Road Network used as test data

3.2 Graph Development and Data Refinement:

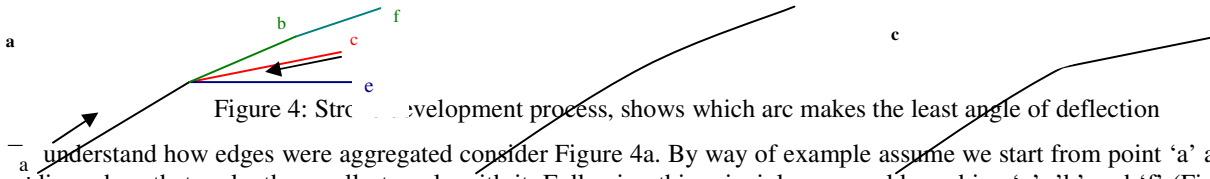
In order to develop a structural representation of the road network a graph was created using the principles of graph theory. Graph theory has been utilized in geographic discipline and especially in map generalisation for quite some time. This is due to the fact that it enables one to characterize topological relationships among objects and has been used to solve such problems such as minimum vertex coloring, modeling Markov chains, critical path analysis, “Bridges of Königsberg” (Mackaness and Mackechnie, 1999). Here a few basic principles are given for complete discussion of Graph Theory see “Pearls in Graph Theory” (Hartsfield and Ringel, 1994).

In simple words a graph is a collection of vertices (nodes) and edges. Vertices are simple objects represents road intersections or ends of roads. An edge is a connection between two vertices or nodes. Edges and Nodes can have attributes such as importance, or name. One of the most important properties of a graph is its ability to model topology explicitly (“the relative location of geographic phenomena independent of their exact position”). To store topology an adjacency matrix is often used – a two dimensional array of a size equal to the number of edges.

A Graph class was implemented in java with Edge and Node as its child classes. Each edge object was allocated an unique identifier, with its start and end nodes along with attributes such as road description and road type were stored in an AAT (Arc Attribute Table) table in Oracle. The unique identifiers were used at the end of the generalisation operations to refer back to the original road segments in the creation of the final output map. An adjacency matrix was developed on the fly to store connectivity information. This was needed for the traversal of the graph in order to maintain connectivity of the output.

3.3 Strokes Development:

Strokes are chains of nodes in which edges that follow the ‘Principle of Good Continuation’ are combined together. Thus edges are aggregated with other edges that make the least angle of deflection with each other as shown in Figure 4.



to understand how edges were aggregated consider Figure 4a. By way of example assume we start from point 'a' and keep adding edges that make the smallest angle with it. Following this principle we would combine 'a', 'b' and 'f' (Figure 4b). The algorithm for aggregation of edges followed the idea by Field et al., (1993) of combining two edges if they make an angle between 40° and 60°. It was also observed that as the size of the stroke increased it was not only the angle of the last edge with the next edge that needed to be considered, but the overall change of the angle from the first edge to the last edge. In algorithm for the stroke development for each selected edge it not only considered the angle between the last edge and new edge but also the angle difference between the first and last edge of the stroke. This ensured that all the strokes were developed following the principle of good continuation.

A further consideration was the direction in which the stroke development starts. As illustrated in Figure 4a assuming if we start from 'a' we select 'b' instead of 'c' or 'e' because both of them make sharper angles than 'b'. Thus we end up with a stroke as shown in Figure 4b. But notice that what constitutes a stroke depends upon the direction in which we start creating the strokes. For example in Figure 4a if we start from 'c' then we end up aggregating 'c' and 'a' together and end up with the stroke as shown in Figure 4c. In such ambiguous cases additional attributes such as the road type were needed so that whenever there was more than one candidate edge the road category was used to select between alternate candidates.

The next step was to classify these strokes using some attribute value. This classification was needed to determine which strokes would be selected for the output map. Any attribute could be used depending upon the application of the resultant map. In this research the length of each edge was used for assignment of weights. So the stroke with longer arcs will have a higher weight than those with a shorter length. The length of all the different strokes was summed. Each individual stroke was divided by this summed length, in order to classify all strokes (Figure 5).



Figure 5: Strokes developed from the graph. The thickness of line is proportional to the calculated weights. The key shows the classification of strokes using their length as basis.

3.4 Principle of Selection:

All strokes cannot be displayed at the output map. Topfer's radical law by Topfer and Pillewizer (1966) was implemented to determine the number of strokes that should be displayed in the resultant map (Dutton, 1999). It should be remembered that this law only gives the number of objects displayed at output scale does not tell us which objects, for this we used the weight to distinguish between important and unimportant strokes. The principle may be expressed in its simplest form as follows:

$$n_f = n_a \sqrt{M_a / M_f}$$

Where n_f is the number of objects which can be shown at the derived scale (in our case number of strokes at output),

n_a is the number of objects (strokes in test data 79) shown on the source material,

M_a is the scale denominator of the source map (for ITN this was assumed to be: 1250),

M_f is the scale denominator of the derived map (in the context of this project: 250,000).

For the test data the number of strokes is

$$\sqrt{M_a / M_f} = 0.07071, n_f = 79 * 0.070711 = 5.585$$

An algorithm was implemented which selected strokes in order of weights until the number of strokes selected are equal to the number of strokes given by the above formula. Using this value and strokes weight; strokes were selected in descending order of weight until number of strokes selected is equal to the value given by formula. The weight of the last stroke selected is the threshold weight for the test data was 0.036.

3.5 Connectivity:

The next step was to remove all the strokes that have weight below the threshold weight as calculated above. The result obtained is shown in Figure 6.

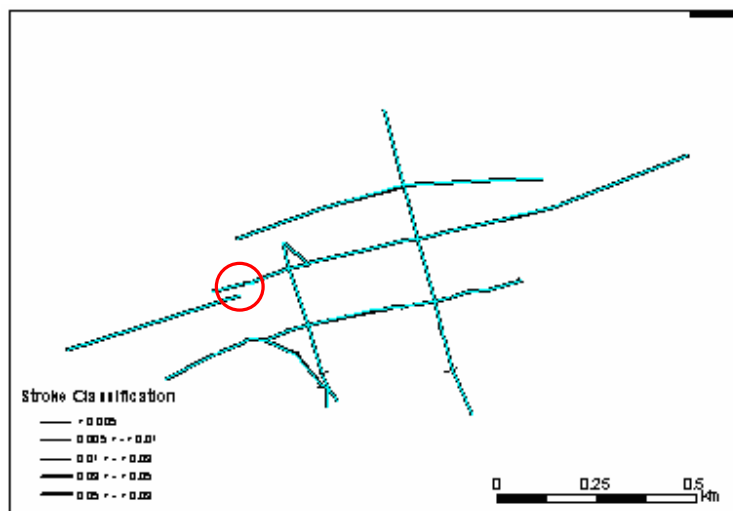


Figure 6: The selected strokes are not connected as shown in the circle.

As illustrated in Figure 7 simple removal of the strokes with smaller weight than the threshold weight is not adequate since it results in disconnection. To maintain the connectivity in the final output the following algorithm was implemented.

Place all the stroke objects in a stack 'S' and set a threshold weight 'W'. Take an empty list 'L', a heap 'H' and a temporary stack 't_s'.

1. Select all the strokes from S which have weight greater or equal to 'W' and place them in t_s.
2. Now arrange the strokes in t_s in ascending order.
3. Next check the connectivity of the strokes. If they are connected then 'end' output is t_s else (not connected) go to 4.
4. Keep removing the stroke from t_s and place them on L and check the connectivity until true' don't understand. Place the connected set of strokes on H.
5. Now take each stroke from the empty list L and find a stroke with the highest weight from S that connects this disconnected stroke with the strokes in H. Store the disconnected and the newly selected stroke in H and repeat 5 until L is empty.
6. The output is H with all strokes connected. You must improve the formatting of material in word.....header bar can be split to control indentation – not carriage return and tabs.

Following the above algorithm for the test data produced the output shown in Figure 7.

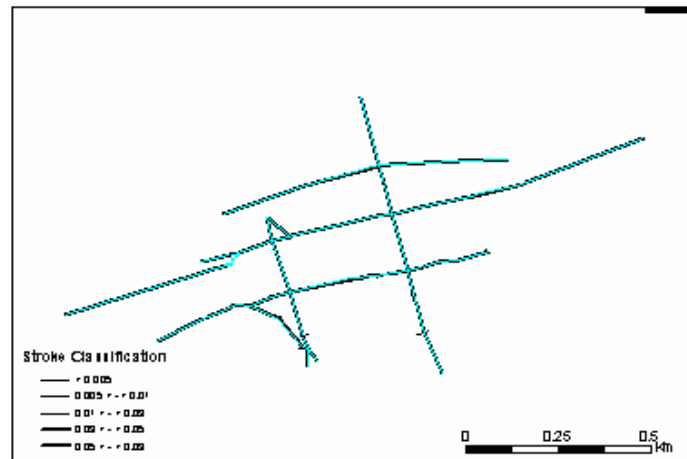


Figure 7: Selected Strokes are connected

3.6 Output:

The last step was to select the 'geographical' description of the roads based on the graph (though the graph remains a very useful basis for further spatial analysis). As discussed above this was simply done by using the unique identifiers for each road segment which correspond to the identifier of each edge object from the database. Thus the final output Figure 8.



Figure 8: Roads selected for the output map

4.0 CASE STUDY:

An important aim of this research was to develop a generalisation process that is applicable to both rural and urban areas. The area selected to check this was 20 * 10 km around The City of Edinburgh. It included both city center and surrounding rural area (Figure 9). This also tested the scalability of the algorithm to see how well it works over large geographical extents, and for mixed urban and rural networks.

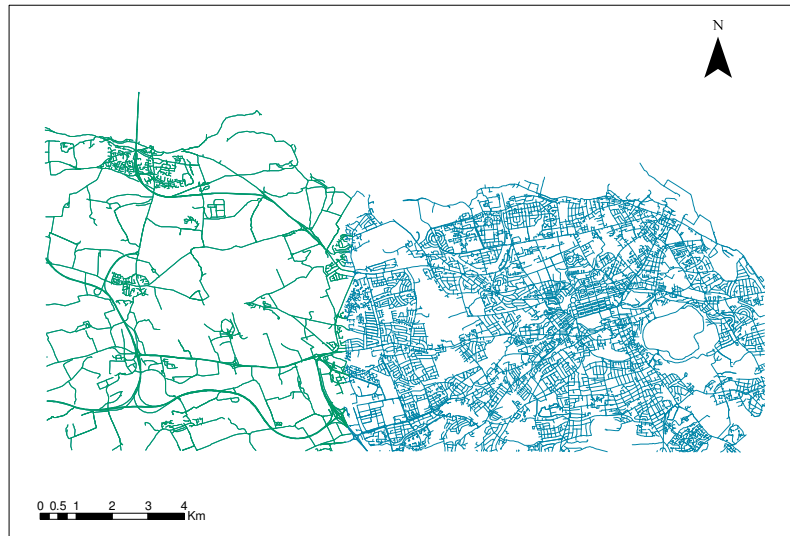


Figure 9: Input Road network at scale of 1:1250 (urban area) to 1:2500 (rural areas)

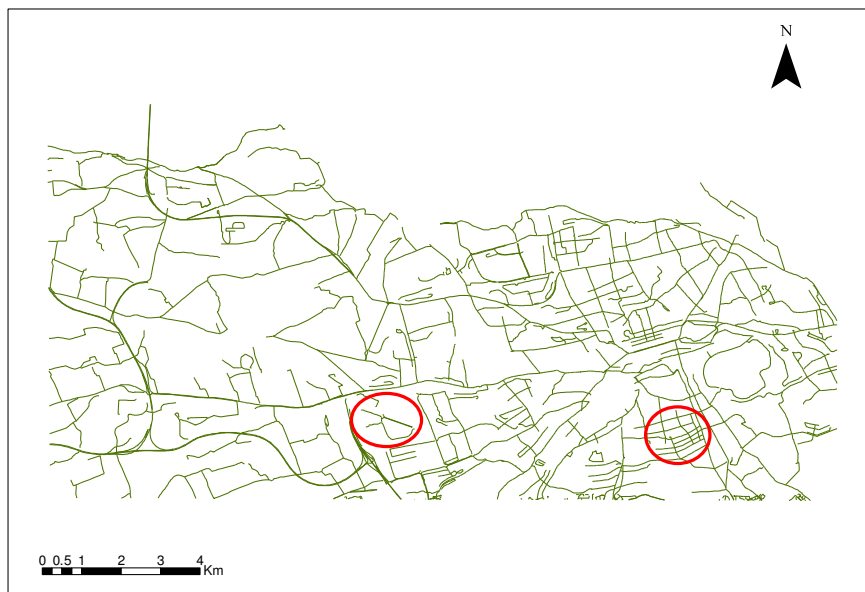


Figure 10: Output selected roads at 1:250,000. The roads in circle are the errors in output.

5.0 EVALUATION:

Evaluation was done firstly by a direct visual comparison between the output (Figure 10) and the OS 1:250,000 Strategi map (Figure 11). Although comparison with paper maps is subjective issue (Mayer, 1998; Weibel and Dutton, 1999) it gives some indication of the success of the algorithm. The results were quite encouraging; it was observed that the algorithm had successfully maintained the overall structure which is same in both maps, although there are a few additional roads selected by the algorithm (marked with circles in Figure 10).



Figure 11:Strategi 1:250,000 by OS

In comparing Figure 10 with Figure 11 it was observed that in most cases the output was exactly the same as the manual one, not only in the overall shape but also the position of roads. In some cases there were additional roads present in the automated output which meant the threshold weight was not correct or there was an error in stroke development. Another error which was quite obvious was the displacement in the OS Strategi dataset which is mainly due to cartographic reasons. Another important issue is explained and illustrated in next section.

5.1 Further Work:

Visually it might appear that a line continues (Figure 12a). Upon detailed inspection, the junction is slightly offset (Figure 12b). In such cases the stroke stops. One way around this problem is through the use of additional attributes together with the angle of deflection. In this way a rule that give precedence to similarity on the basis of attributes such as the type of road over salience can be used but this requires further refinement of the existing algorithm.

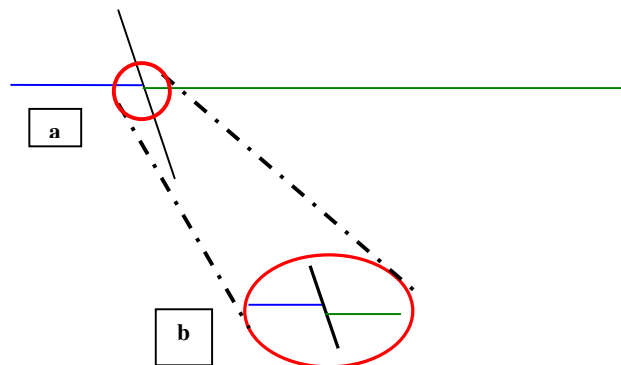


Figure 12: Problem with aggregation using angle measurement

5.2 Utility:

Perceptual grouping principles have been used for generalisation for quite some time, but they have not been tested for the rural and urban areas simultaneously and over large scale changes. The case study presented shows how the roads in the rural areas are connected within the city. The algorithm by default gives more preference to roads outside the city (Figure 10) because of their greater length which is logical so no additional attribute is needed to give these roads more priority. Use of additional attributes is needed only to remove various ambiguities such as the selection of arcs at junctions where there is more than one choice, otherwise the principle worked without the need for additional attributes. It is a useful way of generalisation where the dataset is large and poorly attributed.

5.3 Future Work:

This is by no means a final solution and further enhancement is needed. The perceptual grouping principles allow the delineation of meaningful units of a network in much the same way that the human brain perceives these units. The output provides a framework for generalisation of objects within the bounded set of roads. It will be quite interesting to see how perceptual grouping principles such as similarity can be applied to aerial features including buildings (polygons). The next step would be to check if it's possible to apply the principle of good continuation to dissimilar objects (lines and polygons). Further evaluation can be done via the use of quantification of the output to see if the errors are more common on particular type of roads, and also by buffering the data to see how much it is displaced from the manually generated map.

6.0 Conclusion:

This paper proposed a model based on visual perception for the selection, aggregation and elimination of road segments within a network. Stroke development based on perceptual organization principles provides meaningful results which are similar to the results perceived visually by the human brain. The results showed that simple attributes such as length can be used to determine the order of importance on which the whole process of generalisation can be based. But there is a limit to this principle because removal of less salient features would result in a disconnected network which is not acceptable. The case study showed the success of the perceptual principles applied to rural and urban roads. The implementation is robust enough to deal with large data sets. This paper effectively presented the importance of visual perception techniques in the field of generalisation.

Acknowledgement:

The authors are very grateful for provision of data from the Ordnance Survey via the UK University CHEST agreement.

References:

- Bruce, V., Green, R. P. and Geogreson, A. 2003, '*Visual Perception physiology, psychology, ecology*', 4 edn, Psychology Press New York.
- Chaudhry, O. 2004, '*Perceptual Automated Road Network Generalization over Large Scale Changes*', MSc Dissertation, University of Edinburgh.
- DeLucia, A. A. and Black, R. T. 1987, 'A Comprehensive Approach to Automatic Feature Recognition', in *Proceedings of the 13th International Cartographic Conference*, Morelia, Mexico, pp. 173 - 192 (168-191).
- Dutton, G. 1999, 'Scale, Sinuosity and Point Selection in Digital Line Generalisation', *Cartography and Geographic Information Systems*, vol. 26, no. 1, pp. 33-54.
- Field, T. M., Hayes, A. and Hess, R. F. 1993, 'Contour integration by the human visual system: Evidence for a local "association field".' *Vision Research*, vol. 33, pp. 173-193.
- Hartsfield, N. and Ringel, G. 1994, '*Pearls in Graph Theory: A Comprehensive Introduction*', 2 edn, San Diego, CA: Academic Press.
- Mackaness, W. A. and Mackechnie, G. 1999, 'Automating the Detection and Simplification of Junctions in Road Networks', *GeoInformatica*, vol. 3, no. 2, pp. 185-200.
- Mayer, H. 1998, '*Three Dimensional Generalization of Buildings based on Scale-Spaces*', Report, Chair for Photogrammetry and Remote Sensing, Technische Universität München. Online: [<http://serv.photo.verm.tu-muenchen.de/b3d/s98-html.html>] (1 March 2005)
- Ordnance Survey 2004, 'OS MasterMap Integrated Transport Network Layer', Online: [<http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/itn/>] (1 Sep 2004).
- Sun, M. I. 2004, 'Java.Sun.com The source for Java Technology 2004', Online: [<http://java.sun.com/>] (1 Sep 2004).

- Thomson, R. C. and Richardson, D. E. 1999, 'The 'Good Continuation' Principle of Perceptual Organisation Applied to the Generalisation of Road Networks', in *19th International Cartographic Conference*, ed. C. P. Keller, ICA, Ottawa, pp. 1215-1223.
- Topfer, F. and Pillewizer, W. 1966, 'The principles of selection', *Cartographic Journal*, vol. 3, no. 1, pp. 10 - 16.
- Weibel, R. and Dutton, G. 1999, 'Generalising Spatial Data and Dealing with Multiple Representations', In P. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind, (eds.), *Geographical Information Systems*, vol. and Sons, John Wiley, New York, pp. 125-156.
- Zhang, Q. 2004, 'Road Network Generalization Based on Combined Analysis of Structure and Shape', in *The 11th International Symposium on Spatial Data Handling 2004*, Leciester.